

p53 蛋白质 C 端的三维结构及生物功能区

张彦¹⁾

(昆明医学院生物教研室, 昆明 650031)

刘次全¹⁾

(中国科学院昆明动物研究所, 昆明 650223)

摘要 利用 p53 C 端 118 个氨基酸的 mRNA 二级结构和 Chour-Fasman 蛋白质二级结构预测原则, 预测 p53 蛋白质 C 端 289~325 为卷曲肽段, 368~393 段包括两段螺旋结构: α_1 368~373, α_2 381~388. 其中三段已知的蛋白质二级结构与此 mRNA 二级结构单元间有准确的对应关系. 与四种以多重序列联配为基础的蛋白质二级结构预测方法 (准确率均为 73.20% 左右) 相对照, 预测结果基本一致. 结合单体聚合区 31 个氨基酸晶体结构, 在 SGI INDIGO² 工作站上构建了 p53 C 端 108 个残基的三维结构. 进一步揭示了 p53 C 端诸多生物功能区之间的空间构象关系.

关键词 p53 蛋白质, 二级结构预测, mRNA 二级结构

学科分类号 Q51, Q71

p53 蛋白质 C 端 105 个氨基酸肽段包括了 p53 蛋白质的单体聚合区 (oligomerization), 一个主要和两个次要的核定位区 (NLSs), 以及一个负性调节区 (aa 363~393)^[1]. 在形成 p53 四聚体, 识别并非特异性的结合 DNA 损伤部位, 促使单链 DNA 或 RNA 形成稳定的双链结构, 激活 p53 对 DNA 损伤产生反应, 以及调节 p53 蛋白质核心区与 DNA 的特异性结合等方面都起着十分重要的作用^[2,3].

自 Clore^[4]和 Philip^[5]等发表核磁共振 (319~360) 及 X 射线衍射 (325~365) 的 p53 单体聚合区三维结构研究以来, 迄今尚没有能与核心区三维结构相联系的 p53 C 端完整的三维结构模型. 对 p53 蛋白质 C 端的线性结构和生物功能虽然已有大量分子生物学方面的研究, 但其三维结构的研究迄今仅有单体聚合区的资料. 单体聚合区与核心区的连接部分, 以及单体聚合区 C 端至 p53 蛋白质 C 端部分目前仍然是其三维结构研究的空白区.

在刘次全, 张静等^[6,7]对 72 个已知三维结构的蛋白质样本, 与其编码 mRNA 二级结构的对照性研究基础上 (利用 mRNA 二级结构预测的蛋白质二级结构与实验测定的蛋白质二级结构的符合率达 84.74%. 此外, 对 72 个样本的分析表明, mRNA 二级结构与其编码的蛋白质二级结构之间存在着某些对应关系, 其中包括二级结构单元数的对应, 折叠结构中转角区域的相关性以及结构单元空间排布方式的对应), 我们利用 p53 蛋白质 C 端 118 个氨基酸的 mRNA 二级结构, 对与其对应的蛋白质的

二级结构进行了对比和预测性研究, 并结合 p53 蛋白质单体聚合区 31 个氨基酸序列的三维晶体结构, 尝试性的构建了 p53 C 端 108 个残基序列的三维结构模型.

1 材料及方法

1.1 经计算机互联网从美国国家生物信息中心 (NCBI) GENBANK 数据库中调取 p53 蛋白质序列及其 mRNA 编码序列, 在 IBM486 计算机上用 Zuker 程序 (1992 年, 8.10 版) 折叠与蛋白质 C 端 118 个氨基酸对应的 354 个 mRNA 核苷酸序列, 作出预测的 mRNA 二级结构.

1.2 由美国 Brookhaven 国家实验室 PDB 数据库中调取 p53 C 端单体聚合区的 31 个氨基酸 (336~356) 三维晶体数据, 除此之外的 p53 C 端 289~393 肽段中其他部分蛋白质二级结构, 利用预测的 mRNA 二级结构和 Chour-Fasman 蛋白质二级结构预测原则, 进行蛋白质二级结构预测, 确定 p53 C 端 105 个氨基酸序列的二级结构.

1.3 将 p53 C 端 289~335 和 357~393 两段预测二级结构在 SGI INDIGO² R4400 工作站上分别组装成两段三维结构, 在 Insight II (95.0 版) 程序中的 Amber 力场作初级能量优化, 当其自由能分别为 231 和 163 时 (Cutoff = 1.0 nm, RMS = 4.2 J/mol·nm), 参照 mRNA 二级结构单元间的转折关系, 将这两段三维结构与 PDB 数据库调取的

¹⁾ 云南大学现代生物学中心, 昆明 650091.

收稿日期: 1998-02-24, 修回日期: 1998-06-09

p53 单体聚合区三维结构 (31 个残基, Cutoff= 1.0 nm 时, 自由能为 1719.902 kJ/mol) 连接, 以此为初始构象, 在 Amber 力场中作进一步的能量优化 (约90 000步, Cutoff= 1.0 nm, RMS 偏差为 0.42 J/mol•nm), 结果得到一个初步的、自由能为 - 1780.502 kJ/mol 的 p53 C 端 105 个残基序列的三维结构. 由于计算量太大及设备条件的限制, 在构建蛋白质三维模型时没能考虑溶剂条件.

2 结果和讨论

2.1 p53 蛋白质 C 端 118 个残基 mRNA 及蛋白质的二级结构

为了使预测的 p53 C 端 mRNA 二级结构具有相当的合理性和可靠性, 在确定 mRNA 起始部位时, 决定由一段完整的已知二级结构区开始, 因而选择了 p53 C 端 276A 的 mRNA 编码 GCC 为起始端, 使被折叠的 mRNA 不仅囊括了其编码的 p53 C 端 289 ~ 393 肽段, 还包括了 p53 核心区一段完整的, 由 X 光晶体衍射测定的 α 螺旋结构 (278P~ 286E). 利用 Zuker 程序折叠这段有 354 个核苷酸的 mRNA 序列, 得到了一个编码 p53 蛋白质 C 端 118 个残基 (276A~ 393D) 的 mRNA 二级结构图 (图 1).

这个由 354 个核苷酸组成的 mRNA 二级结构

中, 有五个相对独立的发夹结构. 正如预计的那样, 已知的三段蛋白质二级结构恰好分别定位于 mRNA 二级结构中的前三个发夹区 (图 2). 第一个发夹区包含了由 X 光晶体衍射方法测定的 p53 核心区最大的、由 9 个残基组成的 α 螺旋 (278P~ 286E)^[8]. 第二、第三发夹则相当准确的分别对应着 X 光晶体衍射测定的 p53 蛋白质单体聚合区的 β 片层 (326E ~ 333R) 和 α 螺旋结构 (335R ~ 355A). 并且, 位于 2、3 两个发夹之间的 GGG (Gly334) 形成的锐形转折, 与 X 光晶体衍射测定的该段 β 片层和 α 螺旋段之间由甘氨酸 334 为顶端形成的近乎对折的转角几乎严格对应^[5].

随后的 4、5 两个发夹区目前没有已经测定的蛋白质三维结构可以参考, 但从 mRNA 二级结构单元与蛋白质二级结构单元之间有着非常近似的数量对应关系考虑, 这两个发夹提示, 在 p53 蛋白质 C 端可能还有两个二级结构单元. 结合 Chour-Fasman 蛋白质二级结构预测原则, 对被发夹区 mRNA 编码的蛋白质残基作二级结构倾向性分析, 我们预测, p53 蛋白质 C 端 357K~ 393D 之间可能存在两段 α 螺旋结构, 即 368H ~ 373K (α_1) 和 381K ~ 388E (α_2), 其位置和数目与 mRNA 的两个发夹结构相对应.

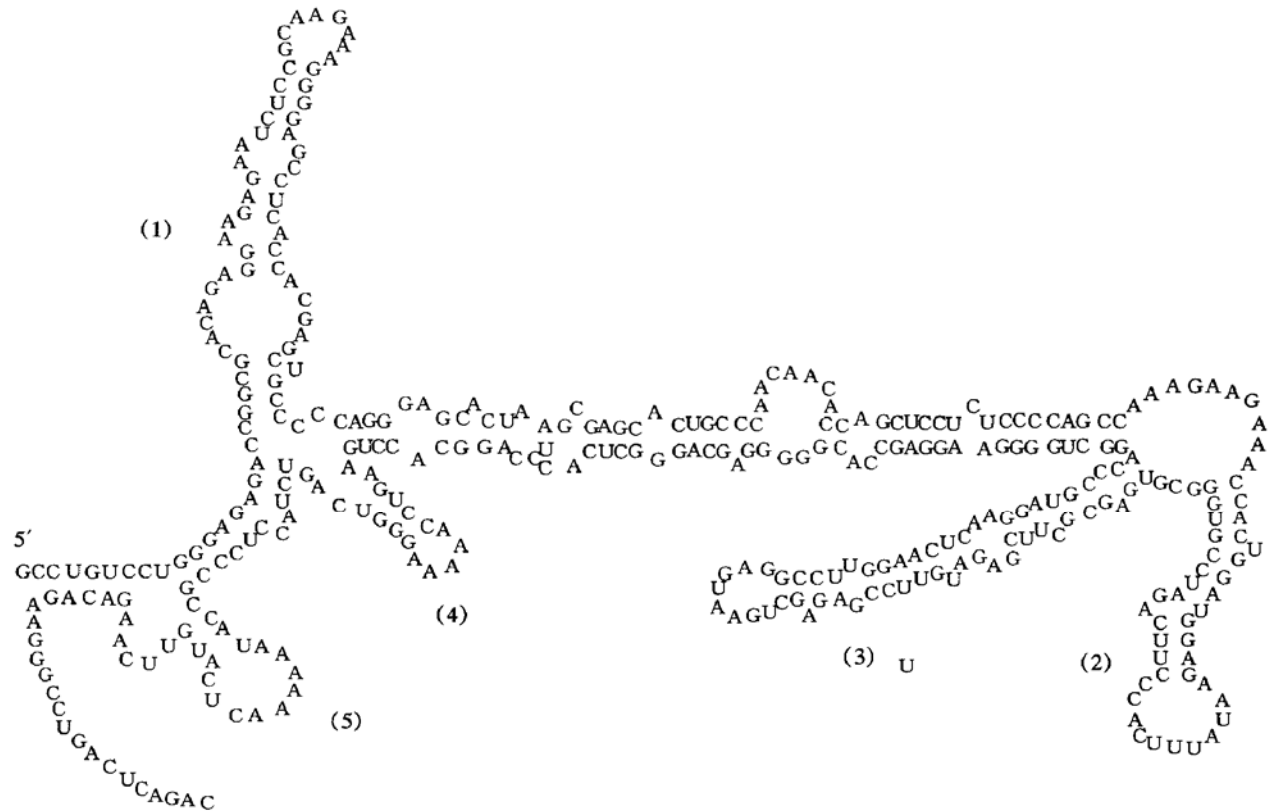


图 1 编码 p53 C 端 118 个氨基酸的 mRNA 二级结构平面图

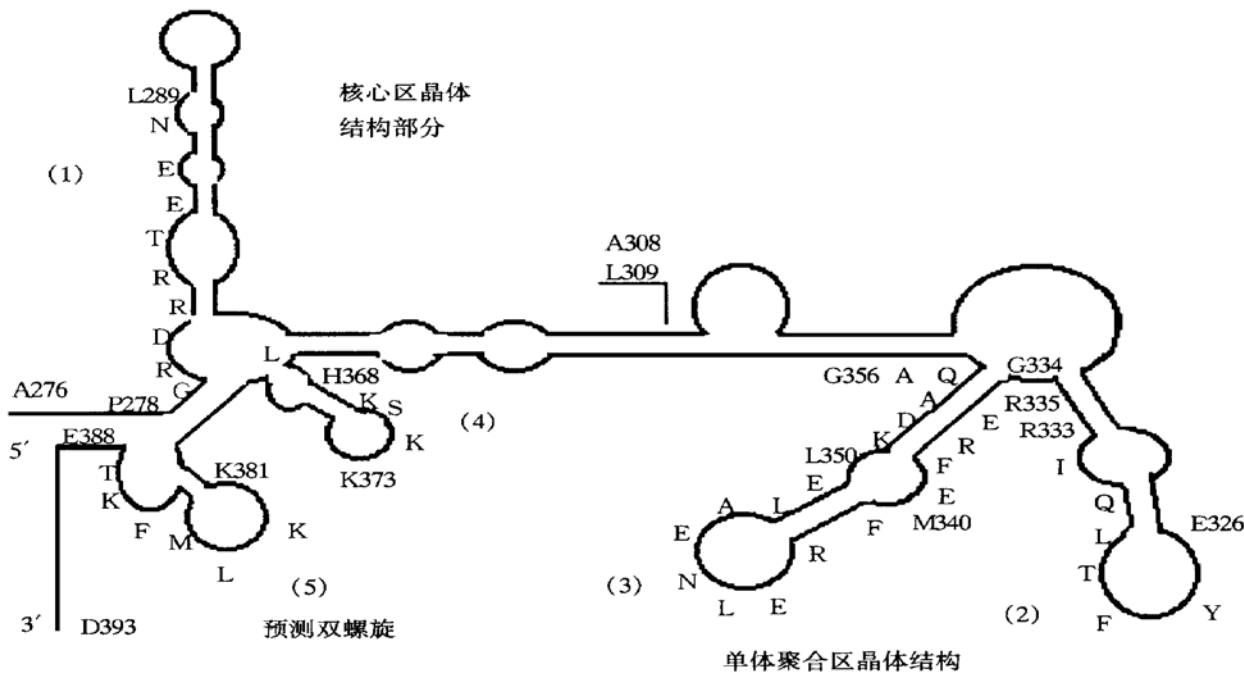


图2 编码 p53 C端 118 个氨基酸的 mRNA 二级结构拓扑图

从残基的组成上看, 预测的这两段 α 螺旋区均为 α 螺旋倾向性很强的氨基酸聚集区. 第一段由一个形成 α 螺旋结构倾向很强的亮氨酸 (L)、三个倾向较强的赖氨酸 (K) 夹一个中立的丝氨酸 (S) 及 N 端的一个强度稍弱的组氨酸 (H) 组成. 第二段由三个 α 螺旋倾向性很强的蛋 (M)、亮 (L)、谷 (E) 氨酸, 三个较强的赖氨酸 (K), 一个较强的苯丙氨酸 (F) 夹一个中立的苏氨酸 (T) 组成. 两段螺旋的 N 端都是较中立的氨基酸 (HSS-, SRH-), 而 C 端则都止于 α 螺旋结构破坏者 (-GQ, -GP). 由于在这两段不大的肽链中都各有三个片层结构破坏者 K, 因而在此形成 β 片层的可能性不大. 组成这两段 α 螺旋结构的残基倾向性强度见表 1.

表 1 组成二段螺旋的残基 α 构象参数及强度¹⁾

氨基酸	E	M	L	K	F	H	T	S
α 构象参数	1.51	1.45	1.21	1.16	1.13	1.00	0.83	0.77
强度	极强	极强	极强	较强	较强	强	中性	中性

¹⁾陶慰孙, 蛋白质分子基础, 北京: 人民教育出版社, 1981, 6.

2.2 p53 蛋白质 C端二级结构预测结果对比

为了将二级结构预测结果的可靠程度作一检验, 并与目前准确率较高的某些蛋白质二级结构预测方法作个比较, 经计算机国际互联网络将 p53 C 端 105 个氨基酸序列送到法国蛋白质生物和化学研究所 (IBCP) 做了以多重序列对准为基础的四种方法的蛋白质二级结构预测, 结果及比较见表 2.

表 2 p53 蛋白质 C 端二级结构预测结果比较

	290	300	310	320	330	340	350	360	370	380	390
p53et	LRKKGEPHHELPPG	STKRALPNNTSSSP	QPKKPLDGEYFTLQ	IRGRERFEMFREL	NEALELKDAQAG	KEPGCSRAHSS	HLKSKKG	QSTSRH	KKLMFKTE	GPDS	D
Gibrat	HHHCCCCCCCCCE	EEEECCCCCCCCC	CCCCCHHHHHHH	CCCCCHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	EEEECCCCCCC	C
Levin	STTTTCTCCCCST	CCCCCTTCCCCCT	TTTCCCCSCCEE	EECCSHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	CCCCCHHCH	EEEECCCCSCC
DPM	CTTCCTCCCTTTT	TTCCHTTTTTCT	CTTCCCTTCC	CEEEHEHCHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	CCCCCTTCT	CTCTCTCHHHHHH
SOPMA	ECCCCCCCCCCCC	CECCCCCCCCC	CCCCCCCCCE	EEEECCCHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	CCCCCCCCC	CCCCCCCCCE
Cons	CTTCCCCCCCCC	CECCCCCCCCC	CCCCCCCCC	EEEECCCHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	HHHHHHHHHHH	CCCCCCCCC	CCCCCCCCC

注: Gibrat, Levin, DPM, SOPMA 分别为四种蛋白质二级结构预测方法, 其准确率都在 73. 20% 左右^[9]. Cons(consensus) 表示这四种方法的综合结果. H: Helix, C: coil, E: Sheet, S: Bend, T: Turn. p53et 栏中: — 表示 X 射线晶体衍射测定的结构. □ 代表用 mRNA 二级结构模板预测的 α 螺旋区.

首先, 四种以序列对准为基础的方法一致确定了在 325G~ 355A 范围内存在两段蛋白质二级结构, 除 Gibrat 方法认为 325~ 332 为 α 螺旋结构外, 其他预测结果与 X 射线晶体衍射测定的一段为 β 片层一段为 α 螺旋结构的情况相吻合, 出入仅在于残基的具体数目和前后位置上的参差. 其次, 这四种预测也提示了在 360~ 390 区间存在两段蛋白质二级结构, 其综合评定结果是在 369L 和 382K 附近有两个小 α 螺旋. 根据该区域在 mRNA 二级结构模板中形成的两个小发夹结构和 Chour-Fasman 的残基倾向原则, 我们预测 368~ 373 和 382~ 388 段为两个 α 螺旋区, 这与对照的四种方法预测结果在蛋白质二级结构的位置、性质方面是一致的, 差别仅在于残基数目的多少. 但真正该区域蛋白质的三维结构如何, 尚有待实验研究的测定和证实. 此外, 289~ 325 段是 p53 蛋白质核心区 (94~ 288) 和单体聚合区 (326~ 356) 两个 X 射线晶体衍射测定结构之间的空白区, 一个主要核定位信号区 (NLS) 亦包括在此间 (316~ 322). 表中四种预测的综合结果认为该区段中 307A、308L 为一小片层结构, 但从 mRNA 模板上看, 该区域为一大段过渡结构, 因而将 289~ 325 段全部作为未定结构 (coil) 考虑.

2.3 p53 C 端的三维结构及其生物功能区

整体上看, p53 C 端 108 个残基的三维构象类

似一个四面柱状体 (图 3, 4), 单体聚合区的片层和螺旋结构分别组成柱状体的两个角, 连接核心区与单体聚合区之间的一大段卷曲肽链 (289~ 325) 和单体聚合区 C 端的 360~ 393 肽段 (包括两段预测的 α 螺旋) 分别占据了柱状体的另外两个角. 由此形成了 p53 蛋白质 C 端空间结构的四个侧面和两个底面: 单体聚合区的片层及其螺旋部分形成侧面一, 单体聚合区的片层与 289~ 325 卷曲肽链构成侧面二, 单体聚合区的螺旋与 360~ 393 肽段形成侧面三, 289~ 325 卷曲肽链与 360~ 393 肽段构成侧面四. 单体聚合区的片层结构的 N 端、螺旋结构的 C 端和 α_2 螺旋组成顶部底面; p53 蛋白质 C 端的 288~ 300 段, 单体聚合区片层与螺旋区的转折部以及 α_1 螺旋形成与核心区连接的底面.

p53 蛋白质线性结构中分别位于单体聚合区两翼的一个主要 NLS (316~ 322) 区及两个次要 NLS 区 (369~ 375, 379~ 384) 在空间结构上它们则相距很近, 分别位于在 p53 C 端整体结构的中央部及侧面. 单体聚合区的片层和螺旋结构形成一面, 便于其与另一 p53 C 端单体聚合区以反平行 β 和反平行 α 的方式形成 p53 二聚体, 进而与其相对的另一二聚体形成四聚体^[5, 10]. 而 MDM2 则通过与这一部位相互作用, 阻碍四聚体的形成, 从而抑制 p53 使变异细胞凋亡的能力. 单体聚合区的 α 螺旋与 $\alpha_1\alpha_2$ 螺旋部分形成的一面, 在识别 DNA 损伤

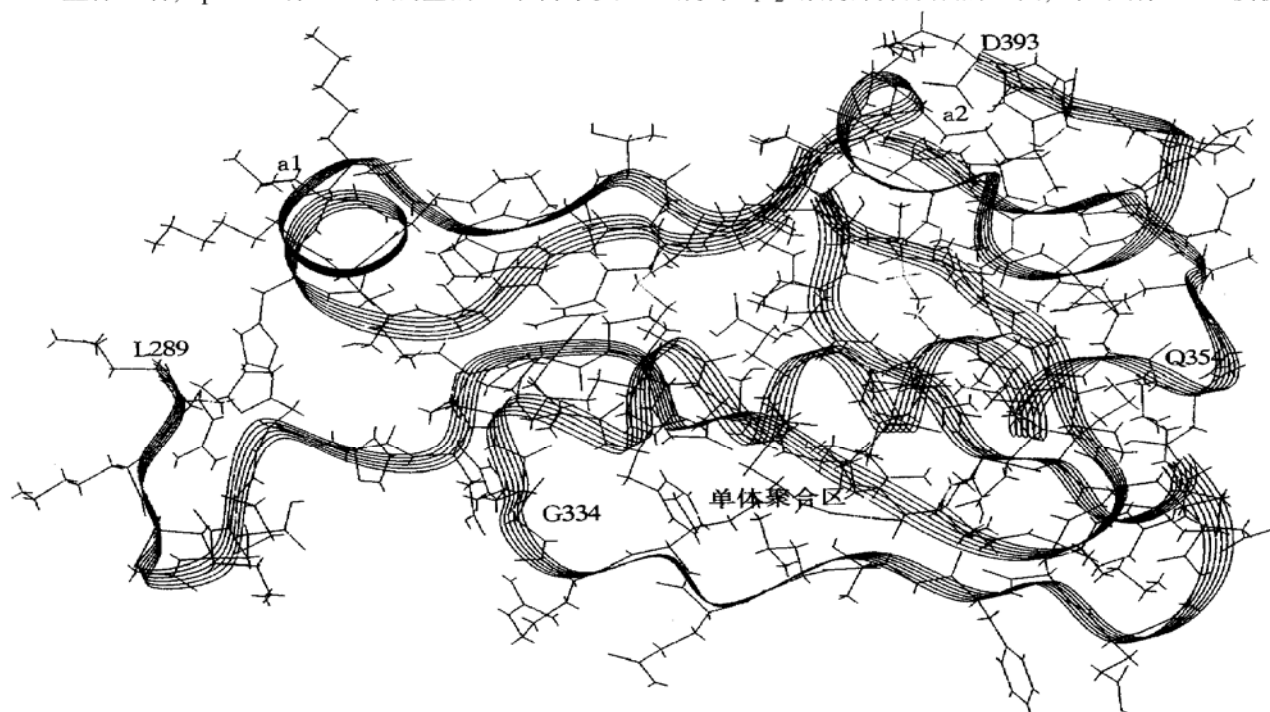


图 3 p53 C 端 108 个残基三维结构侧面

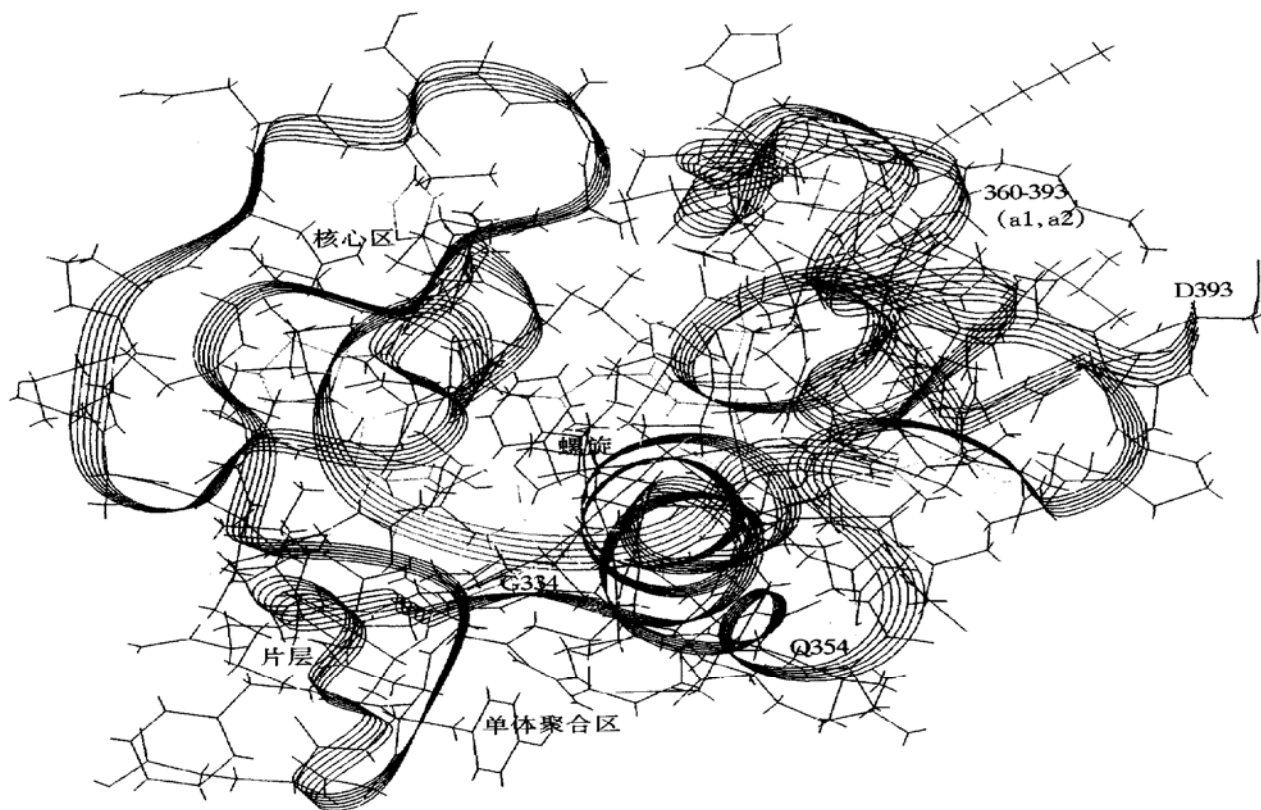


图4 p53 C端 108 个残基三维结构顶面

部位并与之进行非特异结合, 进而激活 P53 分子对 DNA 损伤产生反应过程中起着关键作用^[12, 13]。

大量分子生物学实验表明, 367~ 393 段是 p53 蛋白质非特异性结合 DNA 所必需的, 而 $\alpha_1\alpha_2$ 恰好就在此段并且处在 p53 C 端空间结构表面的一个突出位置。该部与因损伤或试验因素形成的单股核苷酸链结合, 促使其恢复成双股 DNA 或 RNA 链^[14]。免疫分子生物学试验证实, 单抗 Pab421 结合 C 端 370~ 378 段能增强核心区特异性结合 DNA 能力, C 端后 30 个氨基酸缺失也能提高这种能力。从预测的这段三维结构模型来看, 该 C 端以两个小 α 螺旋的形式, 高亲和力、非特异性地与 DNA 结合, 直接起着扫描及识别 DNA 损伤部位, 导介 P₅₃ 特定基因转录激活作用, 以及负性调节核心区 DNA 特异性结合的功能^[15, 16]。

3 结 论

根据 mRNA 二级结构和 Chou-Fasman 氨基酸倾向参数, 预测的 p53 蛋白质 C 端二级结构片段, 与四种以多重序列对准为基础的蛋白质二级结构预测方法相对照, 结果基本一致。预测的 mRNA 二级结构中两个发夹结构与其编码的 p53 蛋白质单体

聚合区两个已知蛋白质二级结构准确对应, 而且两个蛋白质二级结构间的锐性转折与其编码 mRNA 二级结构的两个发夹间的锐性转折也正好准确对应, 这一现象再次表明蛋白质及其 mRNA 的二级结构之间可能存在着某些联系。这种对应关系尚需进一步的探索和论证, 但在目前尚无更好的预测蛋白质二级和三级结构方法可以借鉴的情况下, 我们采用的技术路线和方法或许不失为一种有益的尝试。

参 考 文 献

- 1 Ko L J, Prives C. p53: puzzle and paradigm. *Genes Dev*, 1996, **10**: 1054~ 1072
- 2 Ostermeyer A G, Runko E, Winkfield B, *et al.* Cytoplasmically sequenced wild-type p53 protein in neuroblastoma is relocated to the nucleus by a C-terminal peptide. *Proc Natl Acad Sci USA*, 1996, **93**: 15190~ 15194
- 3 Wu L, Bayle J H, Elenbaas B. Alternatively spliced forms in the carboxy-terminal domain of the p53 protein regulate its ability to promote annealing of complementary single strands of nucleic acids. *Mol Cell Biol*, 1995, **15**: 497~ 504
- 4 Clore G M, Omichinski J G, Sakaguchi K, *et al.* High-resolution structure of the oligomerization domain of p53 by multidimensional NMR. *Science*, 1994, **265**: 386~ 391
- 5 Jeffery P D, Gorina S, Pavletich N P. Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 Angstroms. *Science*, 1995, **267**: 1498~ 1501
- 6 刘次全 (Liu C Q). 信使 RNA 的三维遗传信息研究. 中国科学

- 基金 (Bulletin of National Natural Science Foundation of China), 1998, **12** (1): 23~ 26
- 7 张 静, 刘次全, 来鲁华, 等 (Zhang J, Liu C Q, Lai L H, *et al.*). mRNA 可能携带三维遗传信息. 动物学研究 (Zoological Research), 1997, **18** (2): 封二, 138
- 8 Cho Y, Gorina S, Jeffrey P D, *et al.* Crystal structure of a p53 tumor suppressor DNA complex: understanding tumorigenic mutations. *Science*, 1994, **265**: 346~ 335
- 9 Geourjon C, Deleage G. Significant improvements in protein secondary structure prediction by prediction from mutiple alignments. *Compt Appli Biosc*, 1995, **11**: 681~ 684
- 11 Marston N J, Jenkins J R, Vousden K H. Oligomerisation of full length p53 contributes to the interaction with mdm2 but not PVH E6. *Oncogene*, 1995, **10**: 1709~ 1715
- 12 Lee S, Elenbaas B, Levine A, *et al.* p53 and its 14kDa C-terminal domain recognize primary DNA damage in the form of insertion/deletion mismatches. *Cell*, 1995, **81**: 1013~ 1020
- 13 Bakalkin G, Selivanova G, Yakovieva T, *et al.* p53 binds single-stranded DNA ends through the C-terminal domain and internal DNA segments via the middle domain. *Nucleic Acids Research*, 1995, **23**: 326~ 369
- 14 Hupp T R, Sparks A, Lane D P. Small peptides activate the latent sequence-specific DNA binding function of p53. *Cell*, 1995, **83**: 237~ 245
- 15 Jayaraman L, Prives C. Activation of p53 sequence-specific DNA binding by short single strands of DNA requires the p53 C-terminal. *Cell*, 1995, **81**: 1021~ 1029
- 16 Bayle J H, Elenbaas B, Levien A J. The carboxyl-terminal domain of the p53 protein regulates sequence-specific DNA binding through its nonspecific nucleic acid-binding activity. *Proc Natl Acad Sci USA*, 1995, **92**: 5729~ 5733

The Three Dimensional Structure and Biological Domains of the p53 C-terminus. ZHANG Yan (*Kunming Medical College, Kunming 650031, China*); LIU Ci-Quan (*Kunming Institute of Zoology, The Chinese Academy of Sciences, Kunming 650223, China*).

Abstract It was expected that there are a coil (289 ~ 325) and two α helix (α_1 368~ 373, α_2 381~ 388) structures in p53 protein C-terminal region based on its mRNA secondary structure template and Chour-Fasman's protein secondary structure principle of prediction. The result was conformed by the other four methods of protein secondary structure prediction that are based on the multiple sequence alignment (accuracy = 73.20%). Combine with the 31 amino acids crystal structure of the oligomerization, the three dimensional conformation of p53 C-terminal 108 residues was built using the SGI INDIGO² computer. This structure further expounds the relationship among those biological function domains of p53 C-terminus at three-dimensional level.

Key words p53 protein, secondary structure prediction, mRNA secondary structure

蓖麻毒素 A 链突变体的设计、表达与活性研究

裴武红 冯健男 雷红星 黎 燕 沈倍奋

(军事医学科学院基础医学研究所, 北京 100850)

摘要 利用蛋白质结构同源模建并结合表观静电势分析, 设计了拟具有生物学活性的蓖麻毒素 A 链的突变体. 将 PCR 扩增的突变体基因, 导入 pKK223-3 载体中, 于大肠杆菌 (*E. coli*) 中获得高效、可溶性表达, 而且, 确证了表达产物具有预期的生物学活性.

关键词 蛋白质同源模建, 蓖麻毒素 A 链突变体, 基因表达, 分子设计

学科分类号 R392.11

免疫毒素是由特异性导向的单克隆抗体与强毒性的弹头进行蛋白质交联而得到, 用于治疗肿瘤等顽症. 蓖麻毒素 (ricin) 是常用的弹头成分, 由 A、B 两条链构成, 其毒性部分 A 链是一种 N-糖苷酶, 可以使 28S 核糖体在 4324 位脱去一个腺嘌呤, 从而抑制蛋白质合成, 导致真核细胞的死亡. B 链是一种半乳糖结合型蛋白, 几乎可结合于所有的真核细胞表面. 用原核系统表达重组的蓖麻毒素 A

链 (recombinant ricin toxin A chain, rRTA), 不但可以排除 B 链的影响, 克服了天然 ricin 对细胞的非特异性结合; 而且也不发生糖基化, 在体内不易被快速清除^[1]. 目前免疫毒素在临床上的应用还不理想, 主要原因之一是其分子大, 不易渗透到肿瘤组织内部. 为了解决这一问题, 可以从两个方面着手, 用基因工程方法生产单链抗体代替天然